学术评价视域下单篇论文的学科定位研究*

方胜宇[1]　于　曦[2]
（1. 天津仁爱学院信息与智能工程学院，天津　301636；
2. 天津师范大学图书馆，天津　300387）

摘　要：[目的 / 意义] 梳理学科分类的发展演变，分析比较 CWTS、Citation Topics 和 Dimensions 三个体系对单篇文献的主题定位，为学术评价中以单篇论文的研究主题进行学科定位提供参考。[方法 / 过程] 分析比较三个体系的原理和构建方法，以某高校心理学院的科研论文为研究对象，对其在三个体系中的学科主题分类进行对比分析，并结合论文在 WOS 中的学科分类，分析论文所属学科主题。[结果 / 结论] 三个分类体系各有优势和不足，CWTS 主题分类粒度较细且提供动态图谱，Citation Topics 方便进行主题文献的导出和进一步分析，Dimensions 收录的文献来源和文献类型比较全面和广泛。CWTS 在对主题的描述上要好于 Citation Topics，Citation Topics 对论文学科主题的定位不一定准确，Dimensions 的主题定位粒度相对宽泛一些。对论文研究主题的学科定位需要结合多个学科分类体系，从论文内容、参考文献和施引文献等进行综合判断。

关键词：学科属性　学科分类　主题定位　CWTS　Citation Topics　Dimensions
分类号：G254.2
DOI：10.31193/SSAP.J.ISSN.2096-6695.2023.03.02

1　学科的起源及演变

从人类有系统的生产活动开始，就开展了对学科的探讨。学科的研究可以追溯到公元前 400 年的古希腊时期，柏拉图和亚里士多德都对学科进行了阐释并进行了初步划分。柏拉图认为学科都带有艺术特色，分为“获得性艺术”和“生产性艺术”。“获得性艺术”即不产生任何事物，只是通过言行来征服，或者阻止别人去征服已经存在和已经产生的事物，如数学、政治学和辩证法学科。“生产性艺术”就是使以前不存在的事物存在，如农业、医学等[1]。亚里士多德认为学科

* 本文系天津市哲学社会科学规划（重点）项目“‘双一流’学科建设背景下全域学科分类映射理论与实践”（项目编号：TJTQ21-001）的研究成果之一。
[作者简介] 方胜宇（ORCID：0009-0004-0990-1427），男，讲师，硕士，研究方向为深度学习、自然语言处理，Email：cavalierfang@126.com；于曦（ORCID：0000-0003-4145-1839），女，副研究馆员，硕士，研究方向为学科服务、数据分析，Email：yuxisd_2004@126.com。

就是科学的范畴，包括理论知识、实践知识和生产知识。理论知识包括哲学、数学和自然科学，实践知识指伦理学和政治学，生产知识指艺术可以模仿自然完成的事和艺术完成自然无法完成的事[2]。古罗马时代的马库斯·图利乌斯·西塞罗（Marcus Tullius Cicero）提出学科包括科学和艺术，他认为哲学和口才学是科学，而大部分的自然科学属于艺术[3]。文艺复兴时期的弗朗西斯·培根（Francis Bacon）提出人类的认知可分为历史、诗歌和哲学三大范畴[4]。近现代，工业革命带动科技的进步，随着人们认知的发展，对学科的划分也在不断地发生变化。这时期比较有代表性的学者有让·皮亚杰（Jean Piaget），他认为学科经历了4个演变过程，第一阶段是逻辑和数学，第二阶段是物理学，第三阶段是生物学，第四阶段是心理和生理学，这四个阶段是相互促进的[5]。

进入互联网时代以来，由于在线出版形式的出现，人类产生了大量的科技文献，为了便于管理，人们开始尝试对文献进行学科划分和整理。1957年美国科学信息研究所（Institute for Scientific Information，简称ISI）的尤金·加菲尔德（Eugene Garfield）在美国费城创办了科学引文索引，开展了对文献的科学整理。之后世界各国大学的发展和各大数据库商的纷纷成立也促进了学科的进一步发展和完善，形成了多种多样的学科分类体系，其中的单学科分类体系，如化学的Chemical Abstracts、经济学的Econlit、医学的MeSH；多学科分类体系如科睿唯安公司的WOS学科分类、ESI学科分类和GIPP学科分类，荷兰爱思唯尔公司的Scopus学科分类和EI学科分类。各国根据自身特色和发展也创建了适合自己的学科分类体系。不同的文献类型也有不同的学科分类体系，如针对图书、专利和基金的学科分类。第三方学术评价机构如QS、THE、U.S. News和ARWU也建立了各自的学科分类体系。

2 学术评价中的学科分类

大多数数据库商制定的学科分类体系是以期刊作为基本单位的，不同的数据库有不同的学科分类体系，同一期刊在不同数据库中具有不同的学科定位。以期刊《Spectroscopy and Spectral Analysis（光谱学与光谱分析）》为例，该期刊是中国光学学会会刊，由钢铁研究总院、中国科学院物理研究所、北京大学、清华大学联合承办的学术性刊物，被SCI、EI和北大核心收录。该刊物在不同的数据库中对应不同的学科分类，在ESI中属于“化学学科（Chemistry）”，在WOS中属于“光谱学（Spectroscopy）”，在Scopus、EI和CNKI中归属多个学科分类（表1）。

表1 《Spectroscopy and Spectral Analysis》在不同数据库中的学科（属性）定位

数据库	学科定位（学科属性）
ESI	化学
WOS	光谱学
Scopus	①化学；②物理和天文学；③工程学；④医学
EI	光学、化学、数学统计等60个分类代码
CNKI	①化学；②物理学

高校的学科划分和学院设置与数据库的学科分类也不一致，这就进一步增加了对学术成果进行学科定位的难度。学院的设置与国家的发展和学校的特色相关，而数据库的学科分类是为了方便对文献的管理，两者并无必然的关系。在做科研成果数据统计分析时，我们经常发现某一学科的科研产出来自于多个二级学院的贡献。如表 2 中某高校进入 ESI 材料科学 TOP 1 的科研成果，来自物理、化学、数学等多个学院的贡献。此外由于学科的交叉和融合，学术成果也会涉及到多个学科领域，很难将其归类到某个单一学科。在教育部学科评估中就淡化了学术成果的学科界定，以各高校自己填报为准。在评价时忽略了研究者的学科背景与科研成果学科定位的一致性。

表 2　某高校 ESI 材料科学的学院贡献

学科	二级学院	总发文量（篇）	总被引频次	篇均被引频次
ESI 材料科学	物理与材料科学学院	274	6 287	22.95
	物理与电子信息学院 *	34	551	16.21
	化学学院	140	3 106	22.19
	数学科学学院	3	49	16.33
	生命科学学院	2	46	23.00
	其它学院	15	78	5.20

*“物理与电子信息学院”是原先的学院名称，后拆分为“物理与材料科学学院”和“电子与通讯工程学院”。

学科分类体系的庞杂和混乱，严重影响到了对学术成果的准确定位及进一步的学术评价，进而影响到学科建设、人才引进与学科战略布局。将科学文献划分到适当的学科领域是有效科学计量分析的基本前提之一，在当前国家“破五唯”、提倡代表作评价制度的大环境下，对于如何实现论文的学科定位已经迫在眉睫。由此人们尝试将学科定位的基本单位由期刊过渡到单篇论文，由所属学科过渡到研究主题，通过划分研究主题对论文进行分类。

3　单篇论文的研究主题定位

近年来，关于科学论文主题定位的研究主要分为三个方面：①基于引用的主题定位方法，包括直接引用、共同引用、间接引用和文献耦合，根据算法进行文献聚类，划分研究主题。Klavans 等开展了基于直接引用、书目耦合和共同引用的文献聚类，比较三种聚类方法的主题分类准确性［6］；Ahlgren 等对 PubMed 出版物进行聚类相关性检测的比较研究［7］；魏瑞斌借助主路径和自引网络获取研究主题［8］。②基于文本的主题定位方法，从论文标题、摘要等提取信息，运用机器学习技术将论文自动进行学科归类，如通过 NLP［9-10］、LDA［11-13］、SAO［14］和机器学习模型［15］开展研究主题的识别、前沿预测和演化趋势分析。夏磊通过 AT 主题模型、相似度计算识别学科间交叉主题的研究［16］；杨京等利用 Keygraph 算法提取论文中体现研究主题的关键词，定位论文的研究主题，通过对比研究主题来评价单篇论文的创新性［17］。③基于人工的主题定位方法，即

由作者提供论文的学科归属，如中文文献中作者提供的中图分类号。有学者研究了三种分类方法在论文主题定位准确性上的比较，如 Zhang 等开展了基于引文分类的 Web of Science 学科类别、基于机器学习技术的 Dimensions 的研究领域（FoR）分类和基于作者选择的 Springer Nature 主题分类，对《自然》杂志发表的论文进行分类比较[18]；王欣开展了 InCites Citation Topics 主题分类体系与期刊分类体系间的对比研究[19]；耿海英等对算法构建的论文层次学科分类体系进行了研究综述[20]。现有基于单篇论文的主题定位体系包括莱顿排名的研究领域体系（CWTS fields of science，以下简称“CWTS”）、科睿唯安公司提供的 Incites Citation Topics（以下简称“Citation Topics”）以及 Digital Science 推出的 Dimensions 体系。

3.1　CWTS、Citation Topics 和 Dimensions 体系简介

CWTS 是荷兰莱顿大学科学技术研究中心根据论文的引用与被引关系，通过莱顿算法将论文分成 5 个主要领域和 4 159 个微观领域[21]。5 个主要领域包括：生物医学和健康科学；生命与地球科学；数学和计算机科学；物理科学与工程；社会科学和人文科学。4 159 个微观领域中，对每一个微观领域都会提供：一个领域标识符；该微观领域的出版物数量；该微观领域所属的主要领域；该微观领域发表文章数量最多的 5 种期刊；从微观领域的出版物标题中提取的 5 个特征术语（表 3）。

表 3　CWTS 微观领域组成

领域标识符	出版物数量（2000~2020）	所属主要领域	发文最多的 5 种期刊	特征术语
0	59 354	Biomedical and health sciences	American Journal of Physiology-Heart and Circulatory Physiology	inositol
			Journal of Biological Chemistry	calmodulin
			Journal of Physiology-London	sarcoplasmic reticulum
			European Journal of Pharmacology	trisphosphate receptor
			British Journal of Pharmacology	nitrotyrosine

Citation Topics 是 2020 年 12 月科睿唯安公司在 Incites 数据库中推出的一种学科领域分类方式，2023 年 4 月进行了数据集的更新。Citation Topics 也是基于单篇论文的分类方式，构建宏观、中观和微观主题的三层分类体系，包括 10 个宏观主题、326 个中观主题和 2 437 个微观主题，每个主题标有永久数字前缀和一个特征术语，用以标识精确的主题[22]。例如，微观主题 1.5.77 Deep Brain Stimulation 是中观主题 1.5 Neuroscience 和宏观主题 1 Clinical & Life Sciences 的子主题。

Dimensions 是 2018 年 1 月 Digital Science（数字科学）推出的一个新的科学数据检索和管理平台[23]，该平台收录的文献类型包括期刊、图书、基金、专利、临床试验和政策文件等多种学术资源。使用机器学习和云计算技术，集成和转换数据来创建一致的模型。通过对文献的全文深度索引，包括资助者、研究机构、研究人员或类别状态等内容，为每一篇文献分配相应的学科或研究领域，而不论其来源如何。Dimensions 以澳大利亚 / 新西兰研究领域（Field of Research，以下简称 FOR）作为自己的分类标准。FOR 是澳大利亚和新西兰标准研究分类（Australian and New Zealand Standard Research Classification，ANZSRC）系统的一个组成部分。ANZSRC 用于澳大利亚和新西兰的所有研究和教育领域，它代替了之前的澳大利亚标准研究分类（ASRC），最初于 2008 年 3 月发布，

2019 年对该分类又进行了审查，并于 2020 年 6 月发布了新的分类体系。新分类在内容上更为细化，描述更加清晰。FOR 分类具有三级层次结构：Division、Group 和 Field。Division 代表一个广泛的学科领域或研究范畴，而 Group 和 Field 代表这些类别中越来越详细的子集，共有 23 个 Division、213 个 Group 和 1 967 个 Field。如 Field 300101 Agricultural biotechnology diagnostics（incl. biosensors）属于 Group 3001 Agricultural biotechnology 和 Division 30 Agricultural, veterinary and food sciences。

3.2 CWTS、Citation Topics 和 Dimensions 的原理和构建方法

CWTS、Citation Topics 和 Dimensions 都是基于论文层面的研究主题定位体系，但对于主题定位的原理却并不相同。CWTS、Citation Topics 是根据科学文献之间的关联关系即文献之间的直接引用关系，根据一定的规则制定算法对文献进行自动聚类形成研究主题，通过调整参数，将研究主题聚合成更大的聚类产生学科领域，对研究主题和学科领域赋予标签，最终构建成具有一定层级结构的分类体系[20]。这是一种自下而上的通过文献自组织构建的学科分类体系，在构建的过程中实现了对文献的学科定位，而不是依赖于现有的学科分类体系定位文献主题。Dimensions 是基于现有的学科分类体系，通过机器学习的方法分析文本内容特征，将文献自动归类到学科分类体系中，实现对文献的学科划分。这是一种自上而下的在文献层面上的学科主题定位方法。三个体系的具体构建方法如下：

CWTS 基于对数亿个出版物之间的引用关系的大规模分析，通过算法构建 5 个主要领域和 4 千多个微观科学领域，最新版的 CWTS（2022 年版）将 Web of Science 中自 2000 年至 2021 年的每个出版物（仅限研究论文 Article 和综述论文 Review）分配到 4 159 个领域中的一个，确定 4 159 个微观领域与 Web of Science 中定义的 254 个期刊学科类别（不包括“多学科科学”类别）中的每一个的重叠。Web of Science 中的每个学科类别都与 5 个主要领域之一相关联。基于学科类别和主要领域之间的联系，将 4 159 个微观领域中的每一个分配到 5 个主要领域中的一个或多个。如果微观领域中至少 25% 的出版物属于某一主要领域，则将该微观领域分配给这个主要领域。属于两个主要领域的出版物被分配给这两个领域中的每一个，权重为 0.5。至此，Web of Science 中的每篇论文都会被分配到一个微观领域，而每个微观领域又会分配到一个或多个主要领域[21]。

Citation Topics 将 1980 年至今的 Web of Science 核心合集的所有文献，都根据文献的直接引用关系，通过莱顿算法进行聚类。莱顿算法包括强制聚类和保持最小聚类大小两个聚类参数，1980 年之前出版的文献的引用关系在聚类时也会被考虑到。每一个文献只分配给一个主题，但并不是所有文献都已成功分配给某个主题，从 1980 年起，大约 75% 的文献都分配给了一个主题。超过 90% 的研究论文和综述论文都在聚类计算范围内。对于发表后还没有引用的文献，不参与聚类，也没有分配到相应的主题。Citation Topics 的宏观和中观主题由 ISI 根据其内容手动标记。微观主题被标记为通过算法得出的最重要的关键词。由于引文主题是基于引文关系获得的，而不是组成该主题的文献的主要研究内容，因此提供的主题词可能无法体现出主题中的每篇文献。新文献会根据引用的参考文献添加到现有主题中，并且每个月都会更新数据。每年，都会进行一次完整的聚类更新，在保留引文主题结构的基础上，个别文献可能会在微观主题之间移动，并且可能会出现全新的微观主题。一些微观主题的上层中观主题也会发生改变[22]。

Dimensions 使用现有的分类系统和基于机器学习的方法来自动为所有文献分配一组相关的学

科类别。根据现有关联数据集已建立的研究分类系统，来训练分类算法。使用基于机器学习的逆向工程技术，检查手动编码授权的语料库，由计算机算法生成 FOR 的代码。然后将其与实际代码进行检查，并迭代算法。在 Dimensions 中模拟到分类系统的第二层（Group 层），即 Dimensions 提供的论文学科定位到 Group 层级。除 FOR 代码外，还实施了其他分类系统。这些不同分类系统的选择主要是由研究资助者的需求驱动的，如平台中提供的美国国立卫生研究院的“研究状况和疾病分类系统（RCDC）”和“健康研究分类系统（HRCS）”。为了实现这些方案，已经生成了一种合适的机器学习方法，任何其他分类系统都可以以类似的方式生成，还可以对不属于 Dimensions 中的文档进行分类[24]。

国内关于 CWTS 的研究主要是莱顿世界大学排行榜中大学之间的对比分析[25-26]。关于 Citation Topics 的研究主要是利用 Citation Topics 制定期刊选题策划方案[27-28]。对 Dimensions 的研究是源于 2020 年 7 月“卓越计划”资助的 A、B、C 三类期刊全部集成到 Dimensions 平台上，学者通过 Dimensions 平台对三类期刊影响力和文章特征进行分析[29-30]。目前关于 CWTS、Citation Topics 和 Dimensions 开展论文主题定位的研究文献相对较少。本文对这三种提供主题分类的体系进行对比研究，分析三种体系对论文学科主题定位的特点，以期为单篇论文学科主题的定位分析提供借鉴。

4　三个体系论文主题定位的对比分析

为比较三种体系在定位论文主题时哪个更为准确，本文以某高校心理学院的科研论文为研究对象，分别找到其在 Citation Topics、CWTS 和 Dimensions 中所属的学科主题并展开分析。

4.1　数据来源和获取

本文在 Web of Science 核心合集的 SSCI、SCIE 数据库中检索某高校心理学院 2011~2021 年的论文数据（由于 2022 年的论文不在最新版的 CWTS 统计范围内，因此检索数据不包括 2022 年），检索时间为 2023 年 2 月，文献类型限定为研究论文（Article）和综述论文（Review）。由于高校的合并和拆分，机构名称也在不断地发生变化，加之不规范、甚至错误的书写形式，严重影响了对机构科研产出准确、全面的检索。Web of Science 中关于机构检索主要有三个检索字段标签，即 AD、OG 和 OO。AD 是对机构地址字段的检索，地址字段中包括机构、学院、街道、城市和国家。OG 是对机构首选组织名称（机构的官方名称）的检索，可以检索到机构不同时期拼写形式下的所有记录。OO 只能检索输入的机构拼写形式的记录，该机构其他拼写形式的记录不能被检索到。为保证检索的全面、准确，采用 OG 检索字段检索该高校作为一级机构的全部科研产出，将检索数据导入 DDA（Derwent Data Analyzer）数据分析软件，找到二级机构心理学院的不同书写形式，从而获取心理学院的发文数据，共 269 条。由于这 269 篇论文均出自心理学院，论文的研究主题应都属于心理学范畴。查找这些论文在 CWTS、Citation Topics 和 Dimensions 中的学科定位，分析这些论文的研究主题是否属于心理学范畴，并进一步分析三个体系对论文主题的定位是否准确。Citation Topics 在 2023 年 4 月进行了数据更新，与上一版本相比心理学涉及的中观主题没有变化，微观主题略有差别。但由于笔者下载样本数据和撰写论文的时间是 2023 年 2~3 月，因此样本论文在 Citation Topics 中的主题定位是基于之前的版本。

4.2　三个体系中心理学领域划分

在 Citation Topics 分类中心理学领域归属于“社会科学”宏观主题，包括 2 个中观主题（6.24 Psychiatry & Psychology 和 6.73 Social Psychology）以及 28 个微观主题（图 1）。Dimensions 中对心理学分类归属 Division 52（52 Psychology），包括 6 个 Group 和 36 个 Field（图 2）。



图 1　Citation Topics 心理学涉及的中观主题和微观主题

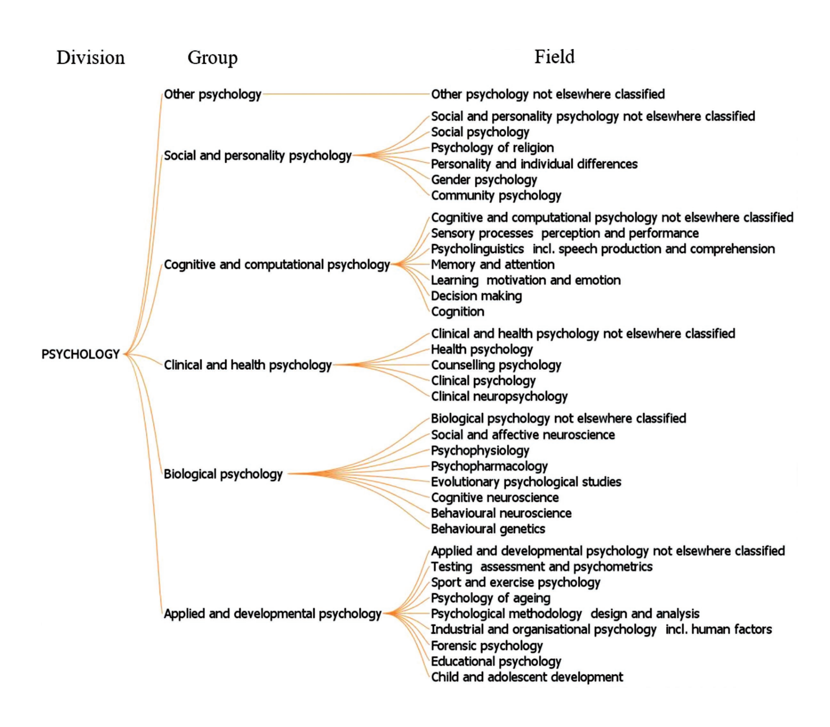


图 2　Dimensions（2022）心理学分类层级结构

CWTS 只提供主要领域和微观领域，主要领域的范围广泛且没有与心理学的直接对应，而微观领域又太过细化。为了查找心理学范畴的微观领域，只能通过每个微观领域提供的发表文章数量最多的 5 种期刊和 5 个特征术语进行人工鉴别。根据汉英大词典提供的“心理”的对应词汇 psychology、mentality、mind、psychic、psycho- 在 CWTS 微观领域中提供的期刊名和特征术语中进行逐一检索判断，找到属于心理学研究范畴的微观领域。在进行人工判断时，我们发现含有以上词语的微观领域不一定属于心理学，如微观领域 1 651 的特征术语中含有“心理距离（psychic distance）”，但是根据该微观领域提供的主要期刊名和其他特征术语，笔者认为这个微观领域是关于“国际商业”的研究，不属于心理学研究范畴，应该排除掉。最后对符合要求的 79 个微观领域认定为心理学范畴。

4.3　论文主题在三个体系中的定位

该心理学院发表的 Web of Science 论文中有 267 篇成功匹配到 Citation Topics 的 22 个中观主题，仅有 2 篇 2020 年发表的论文未匹配成功；有 222 篇成功匹配到 CWTS 的主要领域和微观领域，未匹配成功的基本为 2020 和 2021 年发表的论文；有 266 篇论文可在 Dimensions 中找到研究主题定位，未在 Dimensions 中检索到的 3 篇论文均没有提供 DOI 号。在三个体系中均能成功定位的论文有 217 篇，其中在 Citation Topics 中定位到的最少，仅有 28 篇；CWTS 次之，有 82 篇；Dimensions 最多，有 192 篇。在三个体系中主题定位均属于心理学范畴的论文仅为 16 篇，在两个体系中主题定位属于心理学范畴的论文有 69 篇，仅在一个体系中主题定位属于心理学范畴的论文有 116 篇，在三个体系中主题定位均不属于心理学范畴的论文有 16 篇（图 3）。

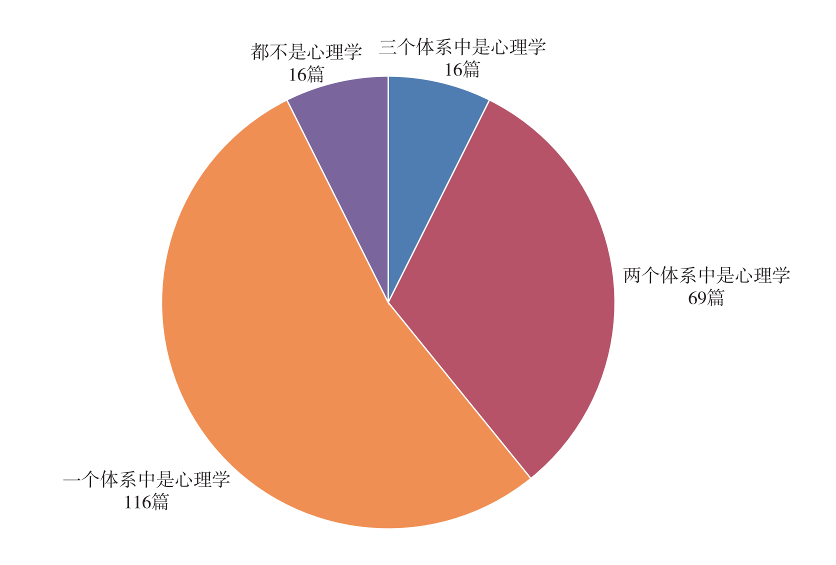


图 3　217 篇论文的心理学分布汇总

4.3.1　三个体系中均属于心理学范畴的论文

Citation Topics 和 CWTS 体系是根据论文的直接引用对论文进行主题定位，因此笔者认为在这两个体系中被认定为心理学的论文其参考文献大部分应该属于心理学范畴的。Dimensions 是基于训练集和机器学习的方法对文献内容特征进行自动分类的，由于训练集和机器学习模型是未知的，因此我们无法通过人为判断分析出这些论文是如何被定位到心理学的。以下仅对在三个体

系中主题定位都属于心理学的 16 篇论文的参考文献进行分析。16 篇论文共有 749 篇参考文献，这些参考文献所属期刊题名中包含 psycho- 的文献有 275 条，也就是说 275 篇参考文献的期刊是明确属于心理学的，除此之外还有一些参考文献的期刊题名中含有“认知（cognition）”“焦虑（anxiety）”“情绪（emotion）”“幸福感（happiness）”“创伤后应激障碍（PTSD）”等心理学术语。对于期刊题名中含有上述词语的参考文献都认定为心理学参考文献，对 16 篇论文的心理学参考文献占比情况进行统计分析（表 4）。

表 4　16 篇论文的参考文献分析

论文	参考文献数	心理学参考文献数	心理学参考文献百分比（%）	论文	参考文献数	心理学参考文献数	心理学参考文献百分比（%）
1	32	29	90.63	9	73	49	67.12
2	76	41	53.95	10	60	33	55.00
3	28	18	64.29	11	34	22	64.71
4	48	36	75.00	12	38	18	47.37
5	53	30	56.60	13	40	17	42.50
6	47	29	61.70	14	42	8	19.05
7	95	73	76.84	15	28	15	53.57
8	73	31	42.47	16	29	17	58.62

16 篇论文的参考文献中仅 1 篇论文的心理学参考文献百分比较低是 19.05%，其他论文的心理学参考文献百分比均超出或接近 50%。

4.3.2　两个或一个体系中属于心理学范畴的论文

在两个体系中主题定位是心理学的论文有 69 篇，其中，同时在 Citation Topics 和 CWTS 体系中属于心理学的论文数为 0 篇，说明这些论文如果在 Citation Topics 和 CWTS 体系中属于心理学范畴，其在 Dimensions 中也会定位到心理学；同时在 Citation Topics 和 Dimensions 体系中属于心理学范畴的论文有 11 篇；同时在 CWTS 和 Dimensions 体系中属于心理学范畴的论文有 58 篇。仅在 Citation Topics 体系中定位为心理学的论文有 1 篇，仅在 CWTS 体系中定位为心理学的论文有 8 篇，仅在 Dimensions 体系中定位为心理学的论文有 107 篇（图 4）。

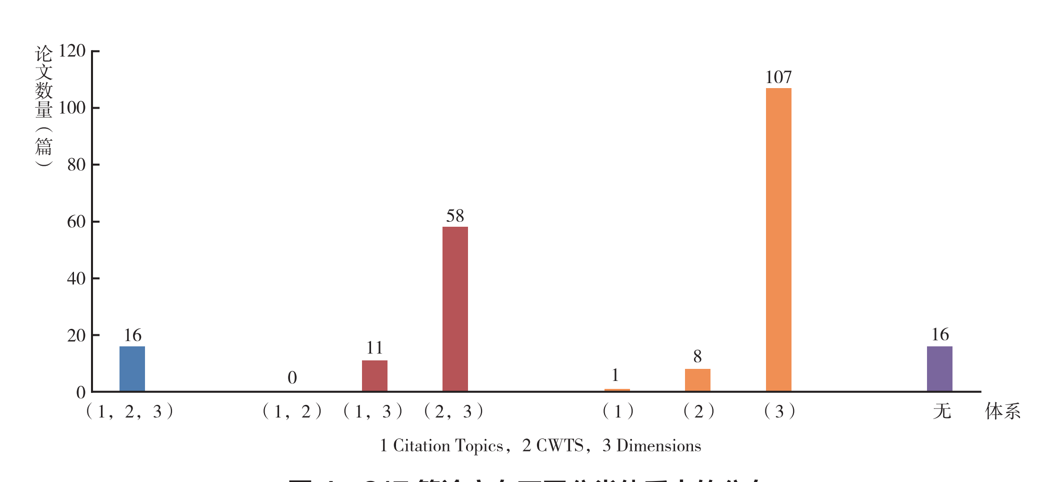


图 4　217 篇论文在不同分类体系中的分布

论文主题定位在三个分类体系中不一致的现象可能与以下原因有关：

① Citation Topics 体系定位到心理学的论文最少，这可能与其包含心理学相关的微观主题较少有关（仅占全部微观主题的 1.15%）。Dimensions 体系中包含心理学相关的微观主题在三个分类体系中是最多的（占全部微观主题的 1.83%），同样 Dimensions 体系中定位到心理学的论文数也是最多的。

② CWTS 没有明确的学科划分，仅提供特征术语和主要期刊，对于学科范畴的归属还需要人工判断，但由于学科背景不同，个人的认知偏差以及相同词语在不用学科中的表述和意义不同，在进行论文学科属性判断时难免会有误差。对于 CWTS 的论文主题定位还需要同行专家进一步验证，以提高论文定位的准确性。

③ Dimensions 定位到心理学的论文数量最多，这一方面是由于其分类来自于现有的学科分类体系，在对论文定位时有非常明确的学科归属，而且新版的学科分类体系中 Group 层级覆盖心理学的范围比较全面，包含 6 个 Group，而旧版的学科分类体系仅包含 2 个 Group。另一方面，Dimensions 对论文的学科定位不是单一学科，被 Dimensions 定位为心理学的 192 篇论文中有 27 篇包含 2 个 Division，即这些论文除了定位为心理学科之外还被定位到其他学科。有些论文即使仅有 1 个 Division，也包括多个 Group。这种学科分配原则大大增加了论文被定位到心理学的可能性。

④心理学的学科特性也是论文在不同体系中定位不一致的原因之一。心理学是一个跨专业性很强的学科，与多个学科都有交叉和联系。对于具有多学科性、多样性的研究领域，学科之间的边界往往相当模糊，将论文分配到适当的学科绝非易事。本文分析的样本中，在一个或两个体系中主题定位为心理学的论文，在其它体系中被定位到的学科主题主要集中在精神病学、神经科学和影像扫描中，以及少量的生物医药、健康科学、教育学、经济学和运动科学等。随着学科的发展和扩大，对于心理学的学科定位更增加了难度。

4.3.3 三个体系中均不属于心理学范畴的论文

在三个体系中均不属于心理学范畴的 16 篇论文中，其学科主要分布在生物科学、临床医学和神经科学方面，其中有 8 篇在三个体系中的主题定位却是一致的。这些论文在 Citation Topics、CWTS、Dimensions 和 WOS（Web of Science 数据库的学科分类）中所属的主题或学科如表 5 所示。

表 5 8 篇论文在 Citation Topics、CWTS、Dimensions 和 WOS 中所属的主题或学科 *

序号	文献标识	学科或主题定位			
	DOI	Citation Topics	CWTS	Dimensions	WOS
1	10.1016/j.nicl.2018.04.004	步态和姿态	肌电图信号，表面肌电图，假手，手势识别，表面肌电信号	医学生理学	神经影像学
2	10.1080/24750158.2019.1702140	信息与图书馆科学	信息素养，公共图书馆，知识组织，图书馆服务，图书馆指导	课程和教育学，教育政策、社会学和哲学，图书馆和信息研究	信息科学和图书馆科学

续表

序号	文献标识	学科或主题定位			
	DOI	Citation Topics	CWTS	Dimensions	WOS
3	10.1080/08946566.2020.1731042	姑息治疗	老年人护理研究	卫生服务与系统，护理学	老年学
4	10.1371/journal.pone.0162876	语言和语言学	声学，声道，声音变化，发音倒置，发声时间	语言研究，语言学	多学科科学
5	10.1044/2019_JSLHR–H–19–0074	语言和语言学	声学，声道，声音变化，发音倒置，发声时间	语言研究，语言学	听力学和语言病理学，语言学，康复学
6	10.3389/fpsyt.2018.00688	神经科学	德拉韦综合征，scn1a，严重肌阵挛性癫痫，kcnq2，通道	临床科学	精神病学
7	10.1186/1744–9081–9–34	神经科学	BDNF val66met 多态性，血清脑，先天性不敏感，神经营养素受体，trka	临床科学	行为科学，神经科学
8	10.3389/fpsyt.2019.00296	神经扫描	托雷特综合征，托雷特，身体变形障碍，强迫症，毛发旺盛症	药理学和制药科学	精神病学

*Citation Topics 列出的是中观主题，CWTS 列出微观领域的 5 个特征术语，Dimensions 列出的是 Group 层次。

论文 1、3 在三个体系的描述中虽然侧重的角度不同，但基本上应属于“医学生理学”（论文 1）和“医学护理学”（论文 3）范畴；论文 2 在三个体系和 WOS 中的描述及分类高度一致都属于“图书馆学”范畴；论文 4、5 在三个体系中的描述也高度一致，归属“语言学”范畴；论文 6、7、8 在 CWTS 中描述的是三种具体的精神疾病，与其它三个体系给出的学科或主题相一致，属于“临床精神病学”范畴。

文中研究的样本数据出自心理学院，样本论文应属心理学范畴。但是样本机构的研究领域已经向医学方向发展，这可能是由于医学中的精神病学与心理学有着千丝万缕的联系。个别研究与心理学范畴相距较远，如语言学、图书馆学。刨除学术不端的可能性，可能的解释是这些研究内容与语言学、图书馆学中涉及到教育学方面的研究有关，如语言教学、二语习得、读者教育、读者心理等。

5 讨论与总结

CWTS、Citation Topics 和 Dimensions 都是基于单篇论文的主题分类，但因其原理和构建方法不同，各有特点。

CWTS 有 4 000 多个微观领域，相较于 Citation Topics 的 2000 多个微观主题，其主题分类更为细化；每个微观领域提供的特征术语有 5 个，对于主题特征的描述更加具体明确；每个微观领域提供发表论文数量最多的 5 种期刊，帮助确定每个主题的核心期刊；提供完整的主题图谱且是动态变化的，可以看到主题随时间的变动。但同时 CWTS 也有一定的缺陷：CWTS 提供的 Web of Science 核心合集中的研究论文和综述论文的主题信息具有一定的时限（仅提供 2000 年到 2021 年出版的论文）；只提供科学图谱和针对单篇论文的主题定位，因此不便于获取某个主题中所有论文的信息，不利于进一步的学科分析；不具有检索功能，没有可供检索下载的专门系统。

Citation Topics 包含 1980 年至今的 Web of Science 核心合集中的所有文献类型；具有检索和下载功能，便于获取某一主题下的全部论文；能快速定位论文所属的研究主题。同样 Citation Topics 也有缺陷：给予论文的主题定位不一定准确，如 6.185 Communication 中的一个微观主题 6.185.1004 Internet Addiction，虽然未归类到心理学中观主题中（6.24 Psychiatry & Psychology，6.73 Social Psychology），但应属于心理学科范畴，目标机构所发表的被归类在 6.185.1004 的论文，其在 CWTS 和 Dimensions 体系中均属于心理学范畴；对于微观主题仅提供一个特征术语，单一特征术语标签标记主题并不科学完善，不能体现该主题的具体内容；不同的微观主题可能具有相同的特征术语，但这些特征术语的侧重点是不同的（如微观领域 3.91.1064 和 3.91.172 的特征术语都是 Heavy Metals，但 3.91.1064 是关于水体沉积物中的重金属的研究，3.91.172 是关于土壤中的重金属的研究，在 Citation Topics2023 年数据更新后，3.91.1064 的特征术语已改为“Sediments”即“沉积物”）；Citation Topics 不像 CWTS 那样提供完整的主题图谱，不能直观地看到学科主题之间的亲疏关系。

Dimensions 检索平台以澳大利亚和新西兰标准研究分类（ANZSRC，2020）中的研究领域（FOR）为学科分类依据，通过基于机器学习的分类方法和构建训练集实现对单个出版物的学科归类。Dimensions 包含的数据源十分广泛，有开放存取期刊（DOAJ）、美国国立医学图书馆数据（PubMed）、巴西科学在线图书馆数据（SciELO）和自然指数期刊数据等。Dimensions 还收录了卓越计划资助的 A、B、C 三类期刊。Dimensions 收录的文献类型丰富，囊括了科学研究生命周期中各阶段的研究成果，在文献类型的覆盖面上远超 CWTS 和 Citation Topics，且对于能提供 DOI 的论文基本都能定位到研究主题，没有 DOI 的论文有时也能定位到研究主题。Dimensions 可通过全文本、题名、摘要以及 DOI 进行单篇论文检索，也允许对某一研究主题的文献进行筛选。除此之外它还提供除传统引文指标之外的两个归一化指标和替代计量指标，为主题文献的深入分析提供参考。虽然 Dimensions 提供了基于论文的学科分类，但自身也存在着一些问题。首先，一些小学科领域没有体现在 Dimensions 的学科研究领域（FOR）中，因此也没有关于小学科论文的学科归属[31]。其次，Dimensions 对文章的学科分类是基于机器学习的分类方法，该方法取决于已建立的分类系统和训练集的大小和质量，训练集的结构和粒度会限制机器学习对单篇论文学科分类的准确性，由于各训练集的大小和质量不同，因此 Dimensions 对文章的学科分类也不完全准确。

再次，Dimensions 对文章的学科定位没有深入到更细化的子主题分类层面（没有定位到 Field 层面），因此对文章的学科定位也不够细化[32]。最后，Dimensions 虽然具有检索和下载功能，但是只能将数据导入 Google Sheets 中，还需要提供机器查询语句才能导出，因此操作对初学者有一定的困难[33]。

三个体系在论文的主题定位上各有优劣势，结合起来使用可以更好地帮助我们分析并定位论文的主题。如果在三个体系中主题定位均一致，就可以认定论文的学科归属。当出现在三个体系中主题定位不一致的现象时要结合多个学科分类体系提供的学科或主题定位，从文章内容本身以及引用关系进行综合判断。基于文章内容的定位方法可通过对题名、摘要的相似性聚类分析生成聚类标签以及通过自然语言处理技术对文章全文提取主题词，根据聚类标签和主题词定位文章的学科（主题）。而基于引用关系的定位方法是分析文章的参考文献和施引文献的学科归属来定位文章的学科（主题）。当然，论文主题定位不一致时也很难确定哪个体系的主题定位是绝对准确的。由于交叉学科、融合学科的现象越来越普遍，有些论文一般会涉及到多个学科领域且这些学科领域之间又有着一定的关系。将学科细分为许多边界明确的领域的想法是理想化的，到目前为止完全令人满意的解决方案是不存在的。但是对于在三个体系中主题定位都与其所在机构的对应学科不一致时，就应该引起科研管理人员和情报分析人员的注意了。

【参考文献】

［1］John M. Cooper. Plato: Complete Works［M］. Indianapolis: Hackett Publishing Company, 1997.

［2］Jonathan Barnes. The Complete Works of Aristotle: The Revised Oxford Translation［M］. Princeton: Princeton University Press, 2014.

［3］Cicero. On the Orator: Book 3. On Fate. Stoic Paradoxes: Divisions of Oratory［M］. Cambridge:Harvard University Press, 1942.

［4］Francis Bacon. Of the Proficience and Advancement of Learning［M］. Cambridge: Da Capo Press, 1970.

［5］Jean Piaget. Classification des disciplines et connexions interdisciplinaires［J］. Revue internationale des sciences sociales, 1964, 16(4): 598-616.

［6］Klavans R, Boyack K W. Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge?［J］. Journal of the Association for Information Science and Technology, 2017, 68(4): 984-998.

［7］Ahlgren P, Chen Y, Colliander C, et al. Enhancing direct citations: A comparison of relatedness measures for community detection in a large set of PubMed publications［J］. Quantitative Science Studies, 2020, 1(2): 714-729.

［8］魏瑞斌 . 基于自引网络和主路径分析的论文主题创新实证研究［J］. 图书情报工作，2018，62（3）：64-70.

［9］蒋彦廷，胡韧奋 . 自然语言处理在其他学科领域的影响考察——基于 CNKI 的中文文献挖掘［J］. 情报杂志，2021，40（12）：169-176.

［10］刘海燕，张志毅，尹晓虎 . 基于论文标题的学科研究主题动力学分析［J］. 情报科学，2019，37（4）：36-43，136.

［11］岳丽欣，周晓英，陈旖旎 . 期刊论文核心研究主题识别及其演化路径可视化方法研究——以我国医疗健康信息领域期刊论文为例［J］. 图书情报工作，2020，64（5）：89-99.

［12］刘博文，白如江，周彦廷，等 . 基金项目数据和论文数据融合视角下科学研究前沿主题识别——以碳纳米管领域为例［J］. 数据分析与知识发现，2019，3（8）：114-122.

［13］吴一平，于纯良，曲佳彬，等．文本主题视域下的高校论文研究前沿领域及演化发展趋势研究［J］．情报科学，2021，39（5）：156–162，183.

［14］马铭，王超，周勇，等．基于语义信息的核心技术主题识别与演化趋势分析方法研究［J］．情报理论与实践，2021，44（9）：106–113.

［15］李欣，温阳，黄鲁成，等．一种基于机器学习的研究前沿识别方法研究［J］．科研管理，2021，42（1）：20–32.

［16］夏磊．基于期刊论文的学科间交叉主题识别研究［J］．新世纪图书馆，2019（12）：62–67.

［17］杨京，王芳，白如江．一种基于研究主题对比的单篇学术论文创新力评价方法［J］．图书情报工作，2018，62（17）：75–83.

［18］Zhang L Sun B Shu F, et al. Comparing paper level classifications across different methods and systems: an investigation of Nature publications［J］. Scientometrics, 2022, 127:7633–7651.

［19］王欣．从学科分类体系到主题分类体系：InCites Citation Topics 主题分类体系解析［D］．大连理工大学，2022.

［20］耿海英，张建东，杨立英，等．算法构建论文层次学科分类体系研究述评［J/OL］．情报理论与实践［2023–5–6］. http://kns. cnki. net/kcms/detail/11. 1762. G3. 20230330. 1028. 002. html.

［21］CWTS Leiden Ranking［EB/OL］.［2022–3–7］. https://www. leidenranking. com/information/fields/.

［22］Introducing Citation Topics in InCites［EB/OL］.［2022–3–7］. https://clarivate. com/blog/introducing–citation–topics/.

［23］一个新的科学数据平台—Dimensions［EB/OL］.［2022–5–29］. http://blog. sciencenet. cn/blog–1792012–1113138. html.

［24］Dimensions Data Guide［EB/OL］.［2022–5–6］. https://www. dimensions. ai/resources/a guide to the dimensions data approach/.

［25］董彦邦．中国大学科研影响力与科研协作力的特点分析——基于 2013 年荷兰莱顿大学排行榜的视角［J］．高教发展与评估，2014，30（5）:19–28，114–115.

［26］苗岩伟，董彦邦．亚洲各国及地区顶尖大学科研影响力与科研协作力的比较分析——基于 2013 年荷兰莱顿大学排行榜的视角［J］．现代教育科学，2014（5）：36–41.

［27］王燕，姚蔚，杜敏，等．利用 InCites 新功能 Citation Topics 助力学术期刊编辑制定选题策划方案——以园艺学科研究领域为例［J］．中国科技期刊研究，2021，32（6）：777–785.

［28］包颖，崔玉洁，文娟，等．基于 InCites 的科技期刊选题策划路径研究——以土壤学为例［J］．西南大学学报（自然科学版），2022，44（12）：221–231.

［29］丁佐奇．基于 Dimensions 平台的卓越行动计划期刊国际影响力评价［J］．中国出版，2020（20）：3–8.

［30］李楚威，丁佐奇．“中国科技期刊卓越行动计划”资助期刊 Altmetrics 评分 Top100 文章特征分析［J］．科技与出版，2020（10）：135–140.

［31］Lutz Bornmann. Field classification of publications in Dimensions: a first case study testing its reliability and validity［J］. Scientometrics, 2018, 117: 637–640.

［32］Christian Herzog, Brian Kierkegaard Lunn. Response to the letter field classification of publications in dimensions: a first case study testing its reliability and validity［J］. Scientometrics, 2018, 117: 641–645.

［33］李洁，孟烨，金佳丽，等．新兴科学引文索引数据库的比较研究［J］．大学图书馆学报，2021，39（6）：48–55，77.

Research on the Discipline Attributes of Literature from the Perspective of Academic Evaluation

Fang Shengyu[1] Yu Xi[2]

(1. School of Information and Intelligence Engineering, Tianjin Renai College, Tianjin 301636, China; 2. Library of Tianjin Normal University, Tianjin 300387, China)

Abstract: ［**Purpose/significance**］To sort out the development and evolution of disciplinary classification, analyze and compare the thematic positioning of individual literature using the CWTS, Citation Topics, and Dimensions systems, and provide reference for academic evaluation using the research topic of individual papers for disciplinary positioning. ［**Method/process**］Analyze and compare the principles and construction methods of the three systems. Taking a research paper from a psychological college of a certain university as the research object, compare and analyze its subject topic classification in the three systems. Combined with the subject classification of the paper in WOS, analyze the subject theme of the paper. ［**Result/conclusion**］Each of the three classification systems has its advantages and disadvantages. CWTS topic classification has a finer granularity and provides dynamic graphs. Citation Topics facilitate the export and further analysis of topic literature. The sources and types of literature included in Dimensions are relatively comprehensive and extensive. CWTS is better at describing topics than Citation Topics, as Citation Topics may not necessarily accurately position the subject topic of the paper. The granularity of topic positioning in Dimensions is relatively broad. The disciplinary positioning of the research topic of the paper needs to be combined with multiple disciplinary classification systems, and a comprehensive judgment should be made from the content, references, and cited literature of the paper.

Keywords: Discipline attributes; Discipline classification; Topic orientation; CWTS; Citation Topics; Dimensions

（本文责编：王秀玲）